

# **SPECIFICATION**

## **TITLE**

5                   **“SYSTEM AND METHOD FOR RECEIVING AND LOADING FARE  
AND SCHEDULE DATA”**

## **BACKGROUND OF THE INVENTION**

10           The present invention relates to an improved system and method for receiving and  
loading updated data sets to a system configured to search such data sets with the purpose of  
identifying data elements within the data set that meet certain user defined search criteria. The  
invention has particular relevance for on-line search systems geared toward providing airline  
flight schedule and fare data to users. The system and method of the present invention may also  
be readily adapted to provide an improved system and method for receiving and loading updated  
15   data sets for searchable systems geared toward other products, services or any other searchable  
data.

20           In the commercial airline industry tickets are often distributed through Global  
Distribution Systems (GDS), sometimes also referred to as Computer Reservations Systems  
(CRS). These are large computerized reservation booking and ticketing systems such as  
Worldspan, Sabre, Galileo, and others. Until recently airline tickets could be only be booked  
either with the air carrier directly, or through an independent travel or ticketing agent. The most  
common method being through an agent.

25           In the typical travel agency office arrangement one or more computer terminals are  
installed in the agency offices. These dedicated terminals are connected directly with one of the  
major GDS systems. A customer of the travel agency generally visits or calls a travel agent  
associated with the firm and tells the agent his or her travel requirements. The specific details  
provided by the customer may include the destination, the travel dates, the desired departure

times, and so forth. The agent then enters the customers request into one of the computer terminals connected to the GDS. The customer's request parameters are entered via a number of complex codes which require a great deal of training for the agent to learn. Once the customer's request is submitted to the GDS, the GDS processes the request and returns a list of flights that  
5 meet the customer's requirements including the fares associated with each flight. However, due to the large amount of computer overhead required to process each customer request, the GDS does not return every possible flight that might meet the customer's requirements. Rather, the GDS provides a limited list with the flights ranked in a specific order based on various rules established by the GDS. Due to the limited scope of the data returned for each request, travel  
10 agents often are able to quote the best possible fares that might be available.

Commerce over the Internet has the potential to greatly affect the way in which airline reservations, booking and ticketing take place. Already there are many online travel agencies through which individuals can search for fares, book flights and purchase tickets from their own personal computer (PC), without ever speaking to or visiting a travel agent. However, while  
15 these "online travel agents" provide reservations booking and ticketing services in a somewhat more convenient manner, they also suffer many of the same drawbacks as the traditional travel agent/GDS arrangement. For example, with most on-line travel agencies available flights and fares are still searched by the GDS. The on-line agent merely receives the user's request in a new manner, namely, over the Internet, but the back end processing remains the same. The on-  
20 line agent processes the request, generates the appropriate codes, and submits the request to the GDS. The GDS in turn supplies a limited list of flights and fares that meet the customer's requirements. The on-line merchant receives the results from the GDS and packages the results for display by the customer's web browser. The raw data itself, the list of available flights

generated by the GDS suffers from the same limitations and is subject to the same display rules governing the display list provided to the travel agent's hard wired terminal in the traditional travel agent/GDS arrangement.

Recently an improved mechanism for searching for flight schedules and fares has been developed. This improved mechanism involves removing the flight schedule and fare searching functions from the GDS and performing them separately from the actual reservation booking and ticketing process. In this system the flight schedule and fare data which the airlines provide to the GDS are also sent to an offline cache memory from which the data may be transferred to various search engines for processing customer travel requests. This system may be distributed over numerous dedicated processors such that processing individual customer requests does not tie up large amounts of expensive computing resources at the expense of other functions that must be carried out simultaneously. Thus, comprehensive searches may be performed that identify every available flight that meets the criteria established in a customer travel request. Furthermore, the results can be ranked in any order independent of the ranking rules of the GDS, and improved display interfaces may also be devised.

However, while the improved remote caching and searching of flight schedule and fare data is an improvement over a purely GDS based system, there have been problems in implementing such a system on a large scale. One of the significant problems in implementing an off-line caching system is that it must be able to receive new data sets containing revised flight schedules and fare data, and seamlessly transition from searching an old data set to searching a new data set. Furthermore, the transition from searching the old data set to searching the new data set must be synchronized with the GDS. Synchronization with the GDS is necessary so that search results from searching the remote cached data will match the flights and

fares that can actually be booked through the GDS. Thus, it is critical that the remote cache process transitions from the old data set to the new data set at the same time the GDS begins accepting bookings with the new data set. New data sets are made available by the airlines on a regular basis, typically three times per day. Thus, a remote cache based searching system must be capable of receiving new data several times a day with no apparent impact on the fare searching and ticketing process. Until now an efficient system and method for implementing such data set updates have not been available.

### **SUMMARY OF THE INVENTION**

The present invention provides both a system and a method for receiving and loading new data sets into a system for searching such data sets in order to return data elements within the data set that meet certain user defined search criteria. The invention is especially well adapted for use in an on-line internet based search system geared toward providing airline flight schedule and fare data to users, wherein the airline flight schedules and fare data are stored separately and are available for searching independently of a Global Distribution System.

According to a first aspect of the invention, a system is provided for searching a data set in order to identify data elements within the data set that meet certain user defined search criteria. The system includes a data search engine server which communicates with a network file system server. The data search engine includes an large on board data storage memory. In a preferred embodiment of the invention the on board data storage memory is a disc-less 2 Gbyte random access memory device. A plurality a data sets are stored on the network file system server in a manner that allows portions of the data sets to be selectively transferred to the data search engine server on an as needed basis. Search queries are sent to the data search engine server from a query distributor. The data search engine server responds to the queries by

identifying data elements within the accessed portions of the data sets that meet the user defined search criteria.

In an embodiment of the invention a system is provided for loading airline fare and schedule data to a low fare search system configured to process travel requests and present flight data, including fare and schedule data, for flights meeting criteria specified in a user request.

The system includes a fare and schedule master database adapted to receive and store airline fare and schedule data for a large number of airline flights. A first file transfer server adapted to receive fare and schedule data from the master database is also provided. The file transfer server includes a flag for indicating whether new data have been received from the master database.

The flag is monitored by a fare load master server. The fare load master server requests a file transfer of the new fare and schedule data from the first file transfer server when the flag indicates that the new fare and schedule data have been received by the master database. Next, a network file system server communicates with the fare load master server. Once the fare load master server has received the new data set, it transfers the new data to the network file system server. The network file system server stores both an old fare and schedule data set and the new fare and schedule data set simultaneously. Finally, the system also includes a low fare search engine server that communicates with the network file system server. The low fare search engine server is configured to simultaneously operate a first search process for processing user queries using the old fare and schedule data set, and a second search process for processing the same user queries using the new fare and schedule data set.

In another aspect of the invention a method is provided for loading a new data set onto a system configured to search a data set for data that meet certain user defined search criteria. The searching system includes a master database that includes a first file transfer server. The

searching system also includes a data loading master server, a network file system server, and a search engine server operable to process search queries on an old data set and on a new data set simultaneously. The inventive method includes receiving a new data set, storing the new data set in the master database and setting a flag indicating that the new data set has arrived. Once the new data set has arrived, the method involves requesting a file transfer from the master database to the data loading master server and transferring the new data set to the data loading master server. The data loading master server subsequently transfers the new data set to the network file system server. Portions of the new data set are then selectively copied to the search engine server. Finally, the search engine server processes identical first and second queries simultaneously in response to each user search request. The first query is processed using portions of the old data set previously copied to the search engine server, and the second query using portions of the new data set copied to the search engine server.

Yet another embodiment the invention provides a method of loading updated fare and schedule data onto a system configured to search such data for available airline flights and fares meeting user defined criteria. In this embodiment the method includes providing a fare and schedule master database adapted to receive and store airline fare and schedule data for a plurality of airline flights. The method also includes providing a file transfer protocol server adapted to receive fare and schedule data from the master database. The FTP server includes a control flag that provides an indication that new data have been received by the master database. Next, the inventive method calls for providing a fare load master server. Among other things, the fare load master server is configured to monitor the control flag. The fare load master server requests a file transfer of a new fare and schedule data set when the control flag indicates that a new fare and schedule data set has been received by the master database. Next, a network file

system server is provided. The network file system server communicates with the fare load master server. The fare load master server is configured to transmit the new fare and schedule data set to the network file system server and the network file system is adapted to store the previous fare and schedule data set along with the new fare and schedule data set. Finally, the inventive method calls for providing a low fare search engine server which communicates with the network file system server. The low fare search engine server is configured to process user queries searching the old fare and schedule data set and the new fare and schedule data set simultaneously.

### **BRIEF DESCRIPTION OF THE FIGURES**

Fig. 1 is a block diagram of a system for searching a data set to identify data elements within the data set that meet certain user defined search criteria and for loading new data sets into the system for searching; and

Figs. 2a-2d are a flowchart showing a method of loading a new data set on a system configured to search a data set for data that meet user defined search criteria.

### **DETAILED DESCRIPTION OF THE INVENTION**

The present invention relates to a system and method for receiving and loading a new data set into a system for searching through such data sets in order to identify elements of a data set that correspond to user defined search criteria. In a preferred embodiment of the invention, the system for searching the data sets is configured to search airline flight schedule and fare data in order to determine the lowest fares available for commercial flights meeting various search criteria established by a potential passenger using the system. In the low fare search embodiment of the invention the data sets searched by the low fare search system include schedule and fare data for a multiplicity of flights between various locations. The user defined criteria for searching the data may include a flight origination location, a destination, a date of travel, and a

time of day. For example, a user may submit a request for flights between Chicago and San Francisco on May 25, departing before 8:00 a.m. After searching the available data set, the search system will return a list of flights for the morning of May 25 between Chicago and San Francisco leaving before 8:00 a.m. The list will include the fare associated with each flight.

5 Prior to describing the system and method of the present invention as applied to a search system configured to search for airline flight schedule and fare data, it should be noted that the present invention may be applied to search systems configured to search other data sets, such as car rental fares and availability, hotel rates, train schedules and fares and the like.

10 A block diagram of a system **100** for searching a data set, including a system for loading new data sets into the search system, is shown in Fig. 1. The system **100** includes a low fare search master database **106**; a fareload master server **110**; a number of network files system servers **112**; a number of low fare search engine servers (LFS servers) **116**; and a number of query distributors (QD) **118**. The fareload master server **110** communicates with a global distribution system (GDS) **122** and the low fare search master database **106** receives data from outside sources such as ATPCO **102** and Innovata **104**. The GDS **122**, ATPCO **102** and Innovata are operated by third parties, and are not, strictly speaking, part of the system **100**. However, the interaction between the fareload master server **110** and the GDS **122**, and ATPCO **102** and Innovata **104** are integral to the method of receiving and loading new data sets when the present invention is practiced on a system for searching airline flight schedules and fares.

20 The architecture of the low fare searching system is as follows. A large number of LFS servers **116** are provided. The LFS servers are organized in racks, each rack containing up to ten LFS servers **116**. Preferably, each LFS server is a Super Micro 6010H 1U Dual CPU stand alone computer operating with 1 GHZ Pentium 3 processors. Each LFS server **116** includes 2



Gbytes of on board random access memory (RAM). A network file system (NFS) server 112 and a query distributor (QD) 118 are provided for each rack of ten LFS servers 116.

Conceptually, an NFS server 112 servicing a rack of ten LFS servers 116 interfaces directly with each LFS server in the rack, as indicated by the connections between the NFS servers 112 and the LFS servers 116 shown in Fig. 1. Physically, however, only a single connection need be provided between an NFS server 112 and its associated rack 114. A switch mounted in the LFS rack 114 distributes the signals between the NFS server 112 and the various LFS servers 116 in the rack 114. Similarly, the QD 118 servicing a rack of ten LFS servers 114 is conceptually connected to each LFS server 116 in a rack 114, but physically only a single signal connection is provided between the QD and the LFS rack 114. A switch distributes the signals passing between the QD 118 and the LFS servers 116.

Each LFS rack 114 along with the NFS server 112 and QD 118 associated therewith operates independently of every other LFS rack 114, NFS server 112 and QD 118. Therefore, the operation of but a single LFS rack 114 and its associated NFS server 112 and QD 118 will be described. The system 100 may be expanded as necessary to include any number of LFS racks 114, LFS servers, NFS servers and QD to meet the demands placed on the system. Each additional LFS rack, LFS server, NFS server and QD will operate in the same manner as the single LFS rack, NFS server and QD described below.

The QD 118 monitors the operation of the LFS servers 116 mounted in the LFS rack 114.

The QD 118 manages the processing work load of each LFS server 116 by distributing search queries among the various LFS servers as they become available. Typically, the query distributor 118 will receive search requests from a user over a computer network. The network may be the Internet, the World Wide Web, or some other communications network. The search

request will include user defined criteria, such as a city or location where a flight is to originate, a destination, and a desired date of travel. Other criteria, such as time of day, airline preference, seating preference and the like may also be included in the search request. The QD 118 monitors the status of each LFS server 116 in the LFS rack 114. Upon receiving a new search request, the QD 118 formulates a search query based on the criteria defined in the search request, and forwards the query to an available LFS server 116 for processing.

After forwarding the query to a specific LFS server process, the QD 118 changes the monitored state of the LFS server process from available to unavailable. The LFS server process remains unavailable until it completes processing the query. Under normal operating conditions, i.e., when a new data set is not being loaded into the system 100, each LFS server 116 is configured to run two search processes at any given time. In other words, under normal operating conditions an LFS server 116 can process two queries simultaneously.

The entire data set through which the LFS servers search for data that meet the criteria established by the user request is stored on the NFS server 112. Preferably, the NFS server 112 is a Compaq DL360 dual processor computer. As an LFS server 116 processes a query, it accesses the data set stored on the NFS server 112 and searches those portions of the data set relevant to the search criteria set forth in the query. As the LFS server 116 searches the data set on the NFS server 112, it simultaneously copies the searched portions of the data set onto its own 2 Gbyte internal RAM. Thereafter, if another query requires a search of the same portions of the data set, the LFS server 116 can search the data from its own internal memory without recourse to the data set stored on the NFS server 112. This arrangement increases the speed with which user requests can be processed, and cuts the power requirements of the system 100 in that fewer LFS servers 116 are required to process a given number of user requests. The power

requirements of the system are also significantly reduced due to the disc-less nature of the LFS servers 16. The 2 Gbyte RAM of each LFS server 116 require approximately half the amount of power consumed by a comparable disc storage system. The disc-less operation of the LFS servers 116 also improves the overall manageability of the system.

5           The system 100 is intended to operate on a continuous basis. Ideally there is to be no down time when users will not be able to access schedule and fare data. In addition to being reliable, the system 100 must be accurate. The data returned in response to user requests must accurately reflect the flight schedules and fares presently offered by the air carriers at the time the user request is filed, so that a user may actually book a flight at the fare provided in response to the user's request. In order for the system 100 to return accurate search results, the data set on which the searches are carried out must be updated from time to time to reflect changes in air carriers' flight schedules and pricing schemes. In light of the continuous operation requirement, updates to the data set must occur without interrupting the process by which user requests are processed. System 100 of the present invention provides a mechanism by which updates to the active data set are provided seamlessly with no noticeable impact on the operation of the system.

Typically, airlines update their fare and schedule data three times per day, although changes may be made more or less frequently as conditions require. Fare data for all commercial air carriers are published by an organization known as the Airline Tariff Publishing Company (ATPCO). Similarly, airline flight schedules are published by the Official Airline Guide (OAG),  
20 or an organization known as Innovata, another organization that publishes airline schedule data in addition to the OAG. Both ATPCO and Innovata provide data in electronic form. The low fare search master database 106, is configured to receive updated fare and schedule data sets from ATPCO and Innovata when the airlines revise their fare structures and schedules. ATPCO

and Innovata transmit the updated data sets to the low fare search master database 106 via File Transfer Protocol (FTP) file transfer. The updated fare and schedule data sets are also sent to the GDS 122. Receipt of the updated fare and schedule data sets initiates the process for loading onto the system 100. In general, the fare and schedule data set loading process is coordinated by the fareload master server 110.

In addition to the system 100, the present invention also encompasses a method by which the seamless updates to new data sets take place. The process for loading a new data set into system 100 is best understood with reference to the flow chart shown in Figs. 2a-2d in conjunction with the block diagram of Fig. 1.

The process begins at step S1. At step S2 the low fare search master database 106 receives and stores updated fare and schedule data set from ATPCO 102 and Innovata 104. Prior to storing the updated data set, the low fare search master database 106 processes the data to convert the raw data into a format compatible with the operating requirements of the system 100. A naming convention has been adopted in the airline ticket distribution industry whereby new data sets may be uniquely identified. According to this naming convention a new data set is given a time stamped file name YYYYMMDD\_\_ (AM1 or PM1 or PM3) where YYYY is the year, mm is the month, and DD is the day that the data set is transferred into the system. The extension AM1, PM1, or PM3 corresponds to the first, second or third data transfer for the given day specified by the file name.

When the low fare search master database 106 has finished processing the received data set, the entire data set is “pushed” to an FTP file server 108 associated with the low fare search master database 106. Two files are created in the FTP file server 108. The first is a large archive file that is made up of a large number of smaller files that make up the actual fare and schedule

data. The archive data set file is identified by the time stamp naming convention described above with a .PGZ file extension appended thereto. The second file is an MD5 checksum file created from the data in the archive file. The data in the MD5 checksum file can be used to verify the integrity of the data in the archive file after the FTP file transfer. The creation of the MD5 checksum files also acts as a control flag indicating that a new data set is ready to be loaded into system 100. The MD5 Checksum file is identified according to the time stamp file naming convention described above, with a \_READY\_TO\_GO.MD5 file extension appended thereto.

At step S3 the fareload master server 110 monitors a directory on the FTP file server 108, looking for the YYYYMMDD\_\_ (AM1 or PM1 or PM3) \_READY\_TO\_GO.MD5 file. At step S4 the fareload master server 110 determines whether a new fare and schedule data set has been received and stored by the low fare search master database 106. If the fareload master server 110 does not locate the file YYYYMMDD\_\_ (AM1 or PM1 or PM3) \_READY\_TO\_GO.MD5 file, it determines that new fare and schedule data have not been received and stored by the low fare search master database 106. In this case, process flow returns to step S3 where the fare load master server 110 continues to monitor the directory on FTP server 108. If, on the other hand, at step S4 the fareload master server 110 detects the YYYYMMDD\_\_ (AM1 or PM1 or PM3) \_READY\_TO\_GO.MD5 file, it determines that a new data set has been received and stored by the low fare search master database 106 and the process moves on to step S5.

At step S5 the fareload master server 110 requests an FTP file transfer from the FTP server 108. When the FTP file transfer is complete, the fareload master serve 110 verifies the integrity of the transferred data at step S7 by comparing the contents of the transferred YYYYMMDD\_\_(AM1, or PM1 or PM2).PGZ file against the checksum data found in the

YYMMDD\_\_ (AM1 or PM1 or PM3) \_READY\_TO\_GO.MD5 file. If an error is detected, the fareload master server may request a second file transfer. At step **S8** the fareload master server extracts the individual files comprising the new data set, and stores the multiple files in a directory named according to the date/time stamp file naming convention. At step **S9** the fareload master server **110** assigns a data set number to the data set according to a 1-6 rolling assignment procedure. A first received data set is assigned data set number 1, the second data set is assigned number 2, and so forth through the number six. The data set numbers are then re-cycled, with the next subsequent data set assigned number 1, and so forth.

At step **S10** the fareload master server **110** sends a command to all of the NFS servers **112** in the system **100**, commanding the NFS servers to pull the new data set from the fareload master server. The command sent to the NFS servers includes the name of the directory where the new data set files are stored, and the data set number (1-6) assigned to the new data set. At step **S11** the NFS servers **112** pull the new data set from the fareload master server **110**. It should be noted that the process of pulling the new data set from the fareload master server **110** to the NFS servers **112** could be reversed with no significant impact on the operation of the system **110**. In other words, the new data set could just as easily be pushed from the fareload master server **110** to the NFS servers **112** as being pulled from by NFS servers. The NFS servers **112** in turn send messages to their associated LFS servers **116** telling them to begin running the two additional search processes on the new data set. The message from the NFS **112** servers includes the number of the new data set.

At this point in the process of uploading the new data set the fareload master server **110** enters a "sleep mode" for approximately 20 minutes while it waits for the GDS **122** to digest the new fair and schedule data. Prior to entering the "sleep mode," however, the fareload master

server **110** sends a message to the QDs **118** alerting the QDs that a new data set is available and identifying the new data set by the assigned data set number.

During this period both the old data set and the new data set are stored on the NFS servers **112**, with each data set identified by its respective data set number (1-6). The QDs send search queries to the two search processes operating on the old data set at step **S13**. The LFS servers **116** process queries searching the old data set at step **S14**. The first pair of LFS search processes continue selectively pulling data from the old data set and searching data from the portions of the old data set previously copied from the NFS servers **112** to the LFS servers' **116** internal memory. Meanwhile, the second pair of LFS server search processes remain idle, selectively pulling only that data from the new data set necessary to initiate the processes and maintain the second processes in a state of readiness so that they will be immediately available when the QDs begin sending queries to the second set of search processes. QDs **118**, which keep track of which search processes are handling which queries and which search processes are operating on which data set, continue to accept the search results from the first pair of search processes being performed by the LFS servers **116** on the old data set until they receive an instruction from the fareload master server **110** commanding them to switch over to the processes searching the new data set.

After its 20 minute "sleep mode," the fareload master server **110** "wakes up," and at step **S15** begins monitoring an FTP file server associated with a Global Distribution System (GDS) **102**, preferably the Worldspan GDS. When configured to monitor the FTP server of the Worldspan GDS, the fareload master server **110** monitors a file directory established by the GDS FTP file server. The GDS receives the updated fare and schedule data set at approximately the same time as the low fare search master database **106**. The GDS must go through its own

internal process of promoting the new data set and demoting the old. The 20 minute delay associated with the fareload master server **110** "sleep mode" allows the GDS **102** to process the new information. It is important that the LFS servers **116** are searching the same data set that the GDS is operating on so that search results returned in response to user queries will match the fares and flight schedules that may be actually booked through the GDS. Thus, the query distributors continue accepting the search results from LFS server search processes operating on the old data set until they receive the signal from the fareload master server **110** indicating that the GDS **102** is up and running with the new data set. The Worldspan GDS provides a signal indicating that its operations have switched over to the new data in the form of a file created in a designated directory of the Worldspan FTP file server. When the Worldspan GDS has completed its promotion of the new data set, it creates a file called SWAP.NOW in the designated directory of the FTP server. The fareload master server **110** monitors the directory looking for the existence of the SWAP.NOW file. If the file exists, the fareload master server **110** knows that the GDS is up and running with the new data set.

At step **S16**, the fareload master server determines whether the GDS is operating on the new data set by determining whether the SWAP.NOW file exists. If the fareload master server determines that the GDS is still running run the old data set, process flow branches off to step **S17** where the QDs **118** continue to accept the search results of the LFS server processes searching the old data set. Thereafter, the searching and monitoring process repeats, as indicated by the process flow returning to step **S13** where the query distributors instruct the LFS servers to continue processing queries with the processes operating on the old data.

If, however, at step **S16** the fareload master server **110** determines that the GDS is operating with the new data set, the fareload master server sends a signal to the query distributors



at step S18 telling the QDs 118 to begin sending queries to the LFS server processes searching the new data set. After a short delay in which searches initiated before the switch command was issued are allowed to complete, a signal is sent from the QDs to “demote” the old data set as shown in step 19. The step of demoting the old data set involves deleting the old data set from the LFS servers’ RAM memory and deleting the old data set from the NFS servers. Upon demoting the old data set, the new data set becomes the de facto old data set as indicated at step S20. The process returns to the beginning at step S21 where it begins again with the next data set update.

Although the present invention has been shown and described with respect to a preferred embodiment thereof, it should be understood by those skilled in the art that the foregoing and various other changes, omissions, and additions in the form and detail thereof may be made therein without departing from the spirit and scope of the present invention.